

Erroneous Requirements: A Linguistic Basis for Their Occurrence and an Approach to Their Reduction

Kimberly S. Hanks

John C. Knight

Elisabeth A. Strunk

University of Virginia



The Problem

- Software is defective
 - Defects cost
 - Some cost more than others
- What we have isn't good enough
 - Too many disasters
 - (and other undesirable events)

Where Are the Defects?

- Lutz result: majority of safety-critical defects derive from poor requirements
- AF Rome Lab result: majority of *all* defects derive from poor requirements
- “The hardest single part of building a software system is deciding precisely what to build.”
-*Brooks*

Role of Domain Knowledge

- The goal of requirements is domain knowledge transfer
 - Requirements is a communication problem
 - Domain knowledge is the thing being communicated
 - Domain experts and developers do the communicating
- “You can learn a lot from the client. Some 70% doesn’t matter, but that 30% will kill you.”
-Paulson

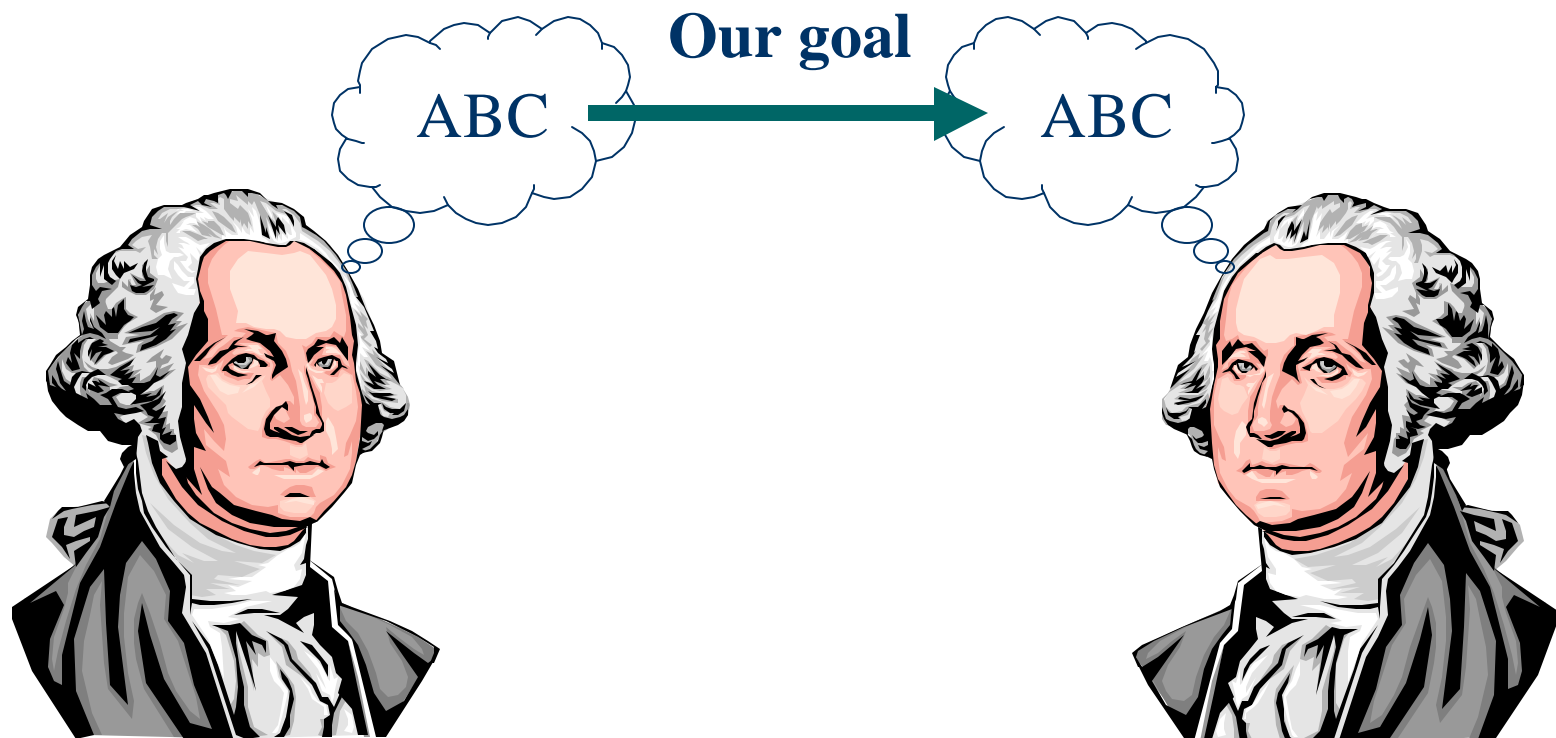
Role of Domain Knowledge

- “It is not the internal complexity of a module but the complexity of the module’s connection to its environment that yields the persistent, safety-related errors seen” -*Lutz*

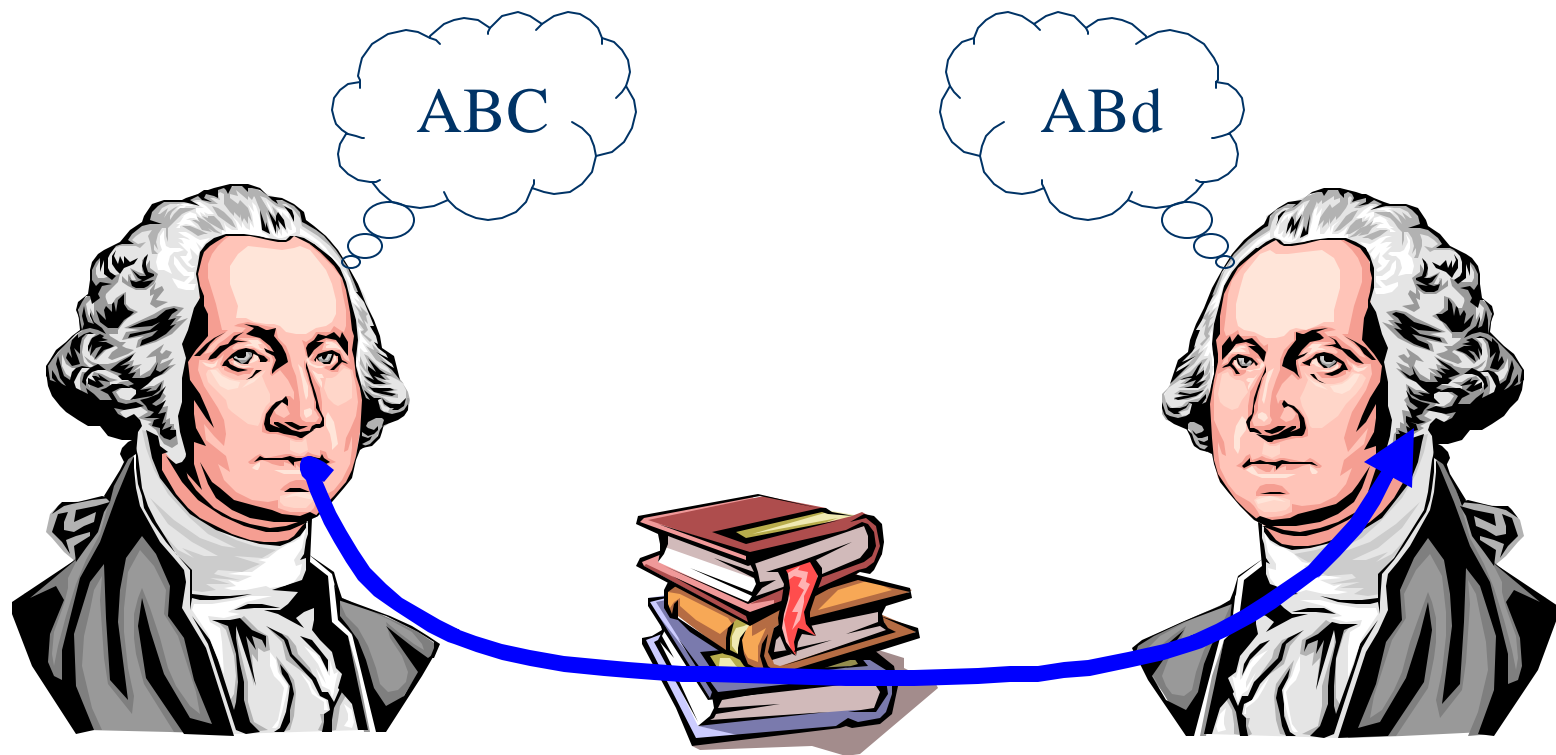
The Communication Factor

- “More has been screwed up on the battlefield and misunderstood at the Pentagon because of the lack of understanding of the English language than any other single factor.”—Vessey

Communicative Goal



Communicative Reality



It's a Communication Problem

- Who studies communication?
 - Linguists
- What do they know?
- Can we use it?

Cognitive Linguistics

- There is a rich theory of linguistics that offers:
 - Explanation of many communication problems that we see in software engineering
 - Structure upon which to build an approach to the use of natural language
- Three major elements:
 - Cognitive categories
 - Cognitive economy
 - Hierarchical structure

Cognitive Categories

- Collections of entities judged to be the same:
 - Example: birds
- Mechanism by which humans store semantics
- Highly structured:
 - Central prototype
 - More and less prototypical members
 - Non members

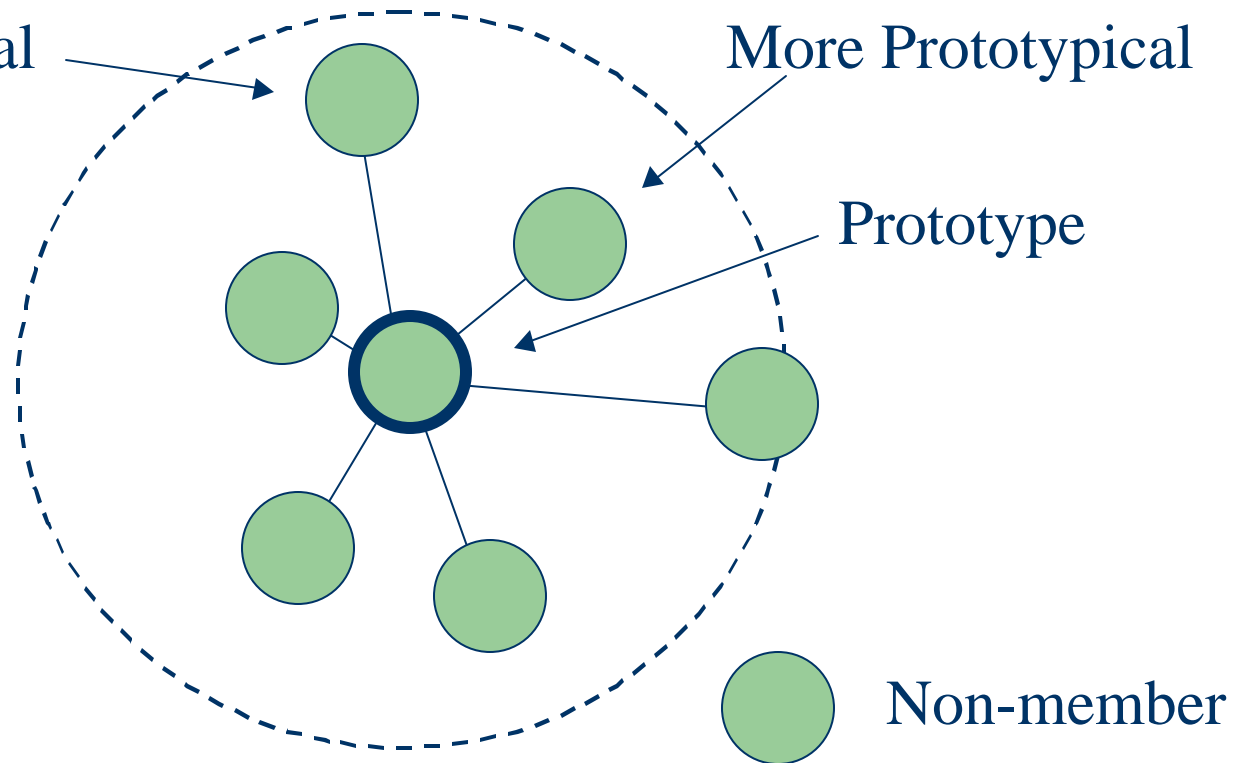
Cognitive Category

Less Prototypical

More Prototypical

Prototype

Sparrow,
Eagle,
Chicken,
Penguin



Cognitive Economy

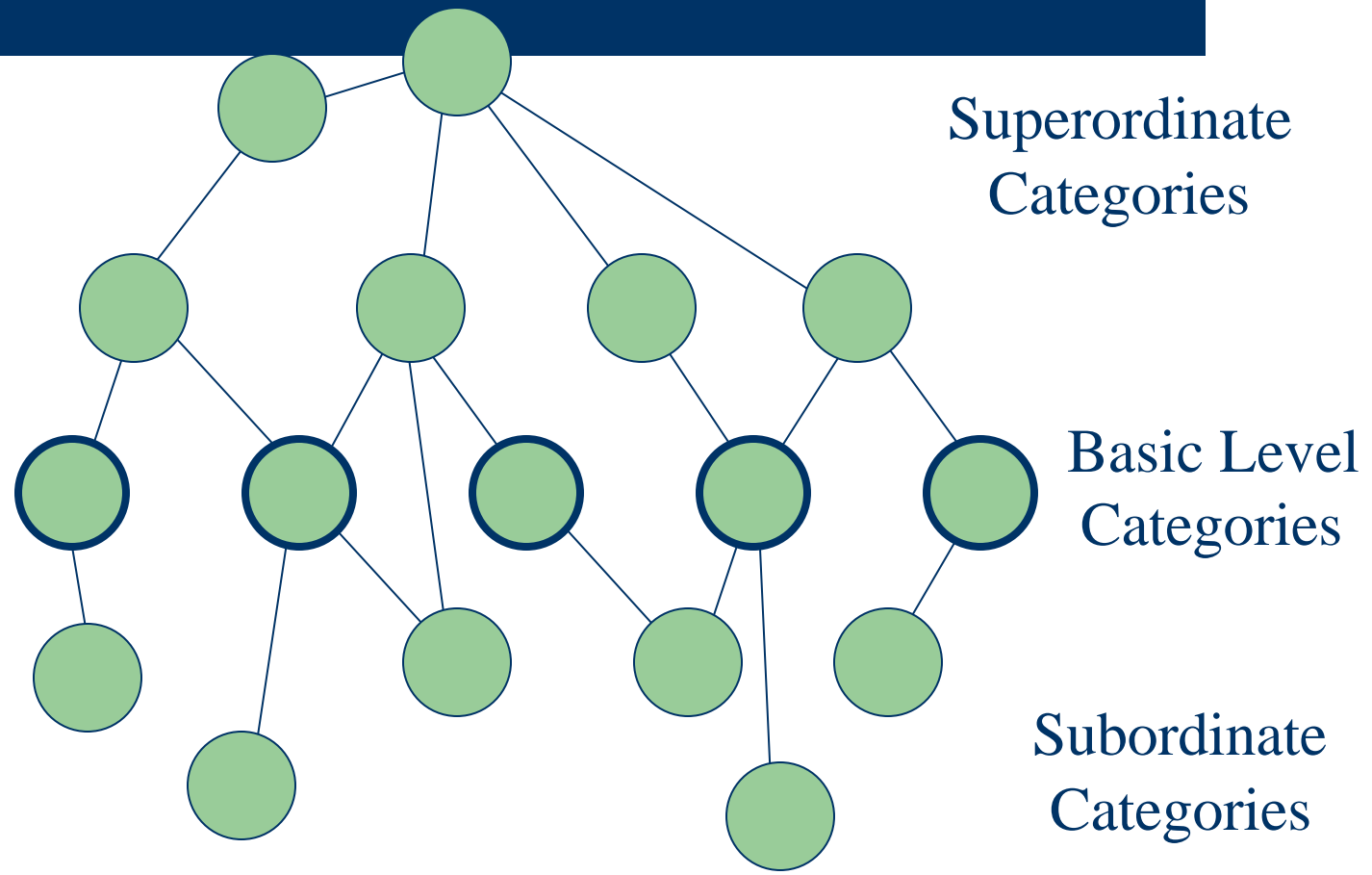
- Attributes assumed based on category name
- What is implied by the word “bird”?
- Problem:
 - Implied attributes different for different people
 - If you know you don’t know, you **ask**
 - If you don’t know you don’t know, you **assume**

This is a very serious source of errors

Hierarchical Structure

- Cognitive categories organized as a hierarchy
- Example:
 - Retriever, dog, mammal
- Basic level in hierarchy has special significance
 - The level most accessed for common communicative needs

Hierarchical Structure



Basic Level Categories

- The basic level is that level at which members of a category have the most in common with each other and the least in common with members of other categories
- Category attributes are the basis
- Experts have been exposed to more attributes for categories within their domain
- Implication:

Basic Level Categories

Experts tend to see lower levels as basic and use them in ways that basic level categories are used.

This translates to proportionally more opportunity for miscommunication.

The Fundamental Problem

Our cognitive machinery is optimized and communication across a domain boundary is not the common case.

It is not a part of human nature to get it right without serious and explicit intervention.

The Domain Map

- An approach to a solution
- Integral part of any and every specification
- Systematic and complete repository of domain terminology
 - Make the implicit explicit
- The only source of reference for domain semantics

The Domain Map

- All domain terms defined
- Semantic relationships made overt
- No cycles
- All definitions grounded in common terms
- Developed by iteration involving:
 - Software engineers
 - Domain experts

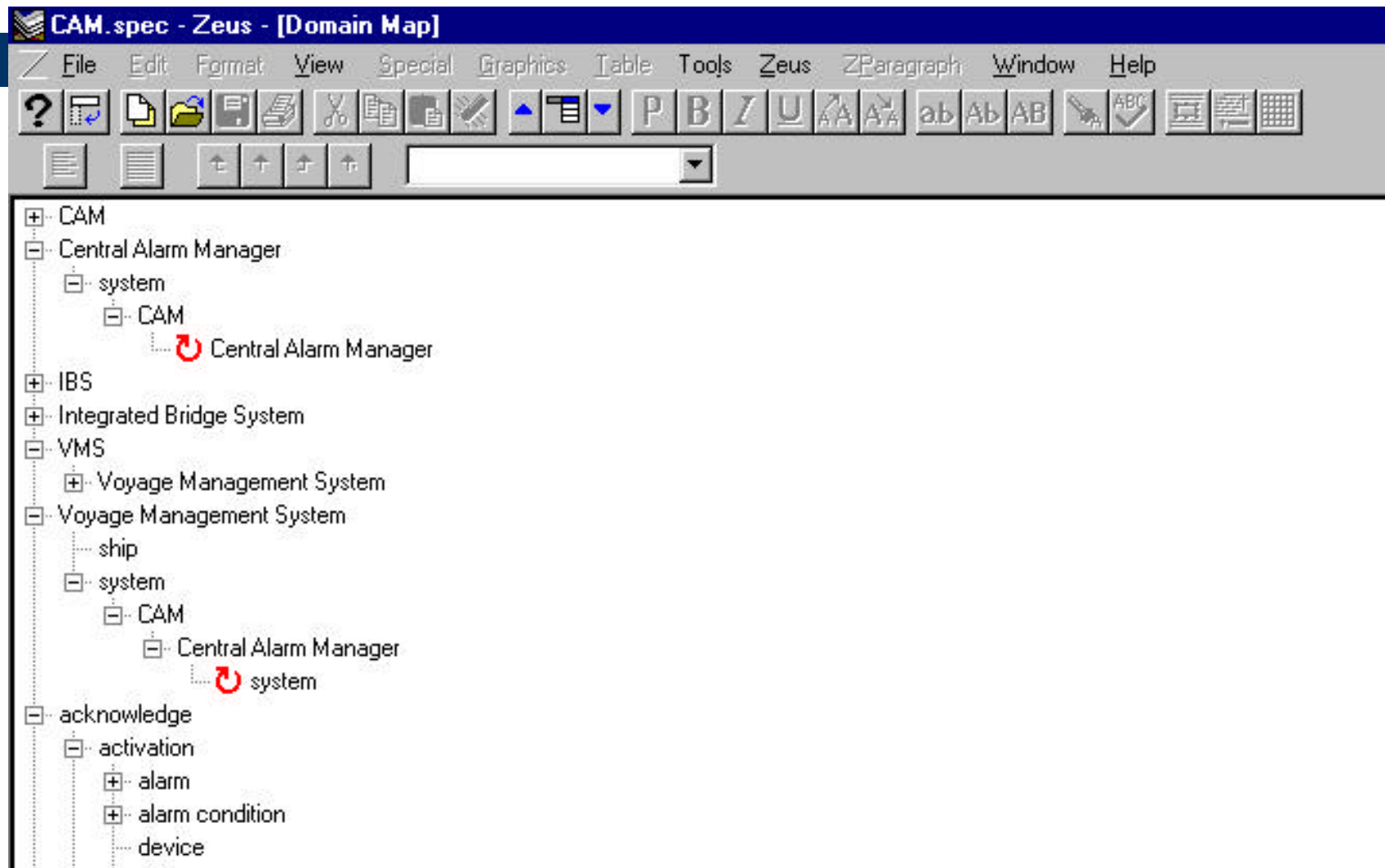
Preliminary Experiment

- Natural language requirements spec for maritime track control system for large ocean-going vessels
- Operator enters waypoints, system steers ship
- 1600 word excerpt, 102 domain terms
- Domain map development:
 - Initial list of terms chosen by developers
 - Reviewed and amended by domain experts
 - Domain map created by developers
 - Reviewed and corrected by domain experts

Preliminary Results

- Initial informal source material:
 - Vague, Incomplete, Inconsistent, Disorganized
- Review of initial map led to 55 corrections (54%)
- Example: “bearing”:
 - Initially understood incorrectly by developer
 - Used by 38 other definitions
- Domain map max height 12, average 5

Zeus Toolset—Reference Tool



Summary

- Requirements is a domain knowledge communication problem
- Linguistics explains a lot
- Our approach consists of recognizing specific linguistic difficulties and systematically removing naturally occurring noise
- Early results from the use of domain maps are promising
- Tool support provides practicality

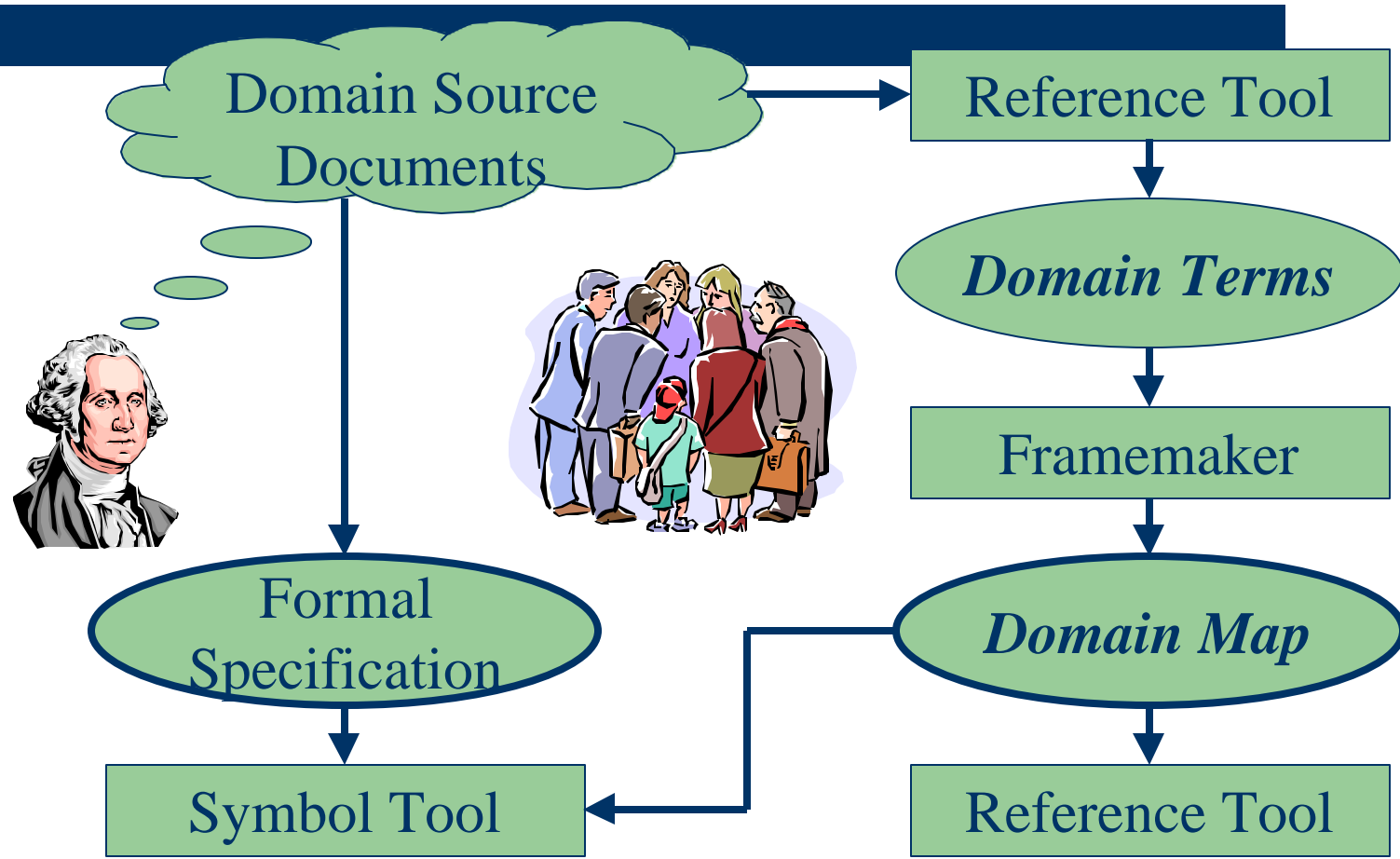
For More Information

- ksh4q@cs.virginia.edu
- knight@cs.virginia.edu
- www.cs.virginia.edu/~jck/recentpapers.htm

Cognitive Categories vs. Types

- Cognitive categories have fuzzy boundaries and graded membership:
 - Is a penguin a bird? (What motivated your answer?)
 - Is it the best (or even a good) example of a bird?
- Formal types have discrete boundaries and flat membership:
 - Is 7 an odd number?
 - Is it any more or less odd than 3? Why?

Tool Support—Zeus



Consequences

- Loss of the Mars Climate Orbiter
 - \$125M for the vehicle alone
- Explosion of chemical reactor
- Plenty of lesser but costly accidents and errors